Surplus Sharing with Coherent Utility Functions

Delia Coculescu [1] and **Freddy Delbaen [2,3],***

[1] Institut für Banking und Finance, Universität Zürich, Plattenstrasse 32, 8032 Zürich, Switzerland; delia.coculescu@bf.uzh.ch
[2] Departement für Mathematik, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland
[3] Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland
* Correspondence: delbaen@math.ethz.ch

Abstract: We use the theory of coherent measures to look at the problem of surplus sharing in an insurance business. The surplus share of an insured is calculated by the surplus premium in the contract. The theory of coherent risk measures and the resulting capital allocation gives a way to divide the surplus between the insured and the capital providers, i.e., the shareholders.

Keywords: coherence; monetary utility; insurance benefit; benefit sharing

1. Notation and Motivation

In the present work, we analyze a method to distribute the surplus of an insurance business between the different agents taking part in the risk exchange. The insured pay premia to an insurance company, and in exchange for this, the company takes over the risks, i.e., the company will pay out the claims that otherwise would have to be covered by the insured. Besides these two players, there is also the supervisor or the regulator. The task of the regulator is to make sure that the insurance business is fair. That means for instance that all companies play with the same rules so that competition can take place. The regulator must also see whether the companies have enough capital to cover the risks, since otherwise—in case of bankruptcy—a substantial portion of the risk would be transferred to society, i.e., to the tax payers or to other economic agents.

As already pointed out by Deprez and Gerber (1985), insurance premia should be dependent on the whole portfolio of insurance contracts. That would mean that the premium to be charged for a contract can only be calculated when all other contracts (signed or in the "pipeline") are known. In practice, this is impossible, and hence, premia are charged that are certainly higher than the fair allocation of the total premium to the individual contracts. The result is that the extra part must be seen as a contribution to the capital of the company and hence is entitled to a share in the eventual surplus.

In the present document, we will deal with a way to calculate the share of each economic agent, be it the capital providers or shareholders or the insured through their extra premium. The problem of how to fix the total amount of regulatory capital is the subject of joint ongoing work, which was initiated by Artzner and Eisele (2018).

We also restrict our analysis to a one-period model. The surplus share is particularly important for life insurance contracts, and handling these would necessitate a multi-period setup. The technical and conceptual problems are not easy, so we "postponed" it to further research.

We will use the language of probability theory, as is usually done in financial and actuarial mathematics. We will fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which all random variables will be defined. In particular, the claims will be seen as random variables defined on this probability space. For simplicity, we will only work with bounded random variables. The vector space of bounded random variables is denoted by $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ or simply L^∞. The restriction to bounded random

variables facilitates the modeling since we do not have to make assumptions on integrability, big tails, and so on. However, it triggers some extra mathematical problems. The solution of these problems will only be sketched, and for more details, we refer to Delbaen (2002, 2011). We assume that the insurance company makes decisions using a coherent utility function.

Definition 1. *A mapping $u\colon L^\infty \to \mathbb{R}$ is called a monetary utility function if the following properties hold:*

1. *if $0 \leq \xi \in L^\infty$, then $u(\xi) \geq 0$;*
2. *u is concave, i.e., for all $\xi, \eta \in L^\infty$, $0 \leq \lambda \leq 1$, we have $u(\lambda \xi + (1 - \lambda)\eta) \geq \lambda u(\xi) + (1 - \lambda)u(\eta)$;*
3. *for $a \in \mathbb{R}$ and $\xi \in L^\infty$, $u(\xi + a) = u(\xi) + a$;*
4. *if $\xi_n \downarrow \xi$ (with $\xi_n \in L^\infty$), then $u(\xi_n) \to u(\xi)$.*

If moreover for all $0 \leq \lambda \in \mathbb{R}$, $u(\lambda \xi) = \lambda u(\xi)$, we call u coherent.

The number $u(\xi)$ can be seen as a risk-adjusted valuation of the future uncertain position ξ. Property (1) in the definition is therefore clear. Risk adverseness is usually translated by concavity properties, and it is believed that combinations are less risky than individual positions. This explains Property (2). Property (3) means that risk-adjusted valuations are measured in money units. Of course, money today (date of the valuation) is different from money at the end of the period. Introducing a deflator or discounting—as has been the practice in actuarial business for hundreds of years—solves this problem. It complicates notation, and as long as there is only one currency involved, it does not lead to confusion if one supposes that this discounting is already incorporated in the variables. The fourth property is a continuity property. Using monotonicity (a consequence of the previous properties; see Delbaen 2011), we can also require that $u(\xi_n) \downarrow u(\xi)$. In this text, we will use the stronger property where $\downarrow$'s are replaced by $\uparrow$'s; see Delbaen (2011). This avoids some mathematical problems that are easily overcome, but they obscure the philosophy of the approach. The homogeneity property is a strong property. Later in this paper, where we use commonotonicity, positive homogeneity is already satisfied.

Sometimes, the value $u(\xi)$ only depends on the distribution or law of the random variable ξ. In this case, we say "law invariant", "law determined", etc. An example of such a mapping is the distorted probability. These were introduced in insurance by Yaari (1987), Denneberg (1989), and Denneberg and Kaplan (1998). They were later used by Wang (2002). An example of such a utility function is (for $\xi \geq 0$) the Choquet integral $\int_0^\infty f(\mathbb{P}[\xi > t])\,dt$ where $f : [0,1] \to [0,1]$ is convex, $f(0) = 0$, $f(1) = 1$. See Delbaen (2011, 1974), Shapley (1972), and Schmeidler (1986) for more details and for the relations between convex games and commonotonicity. See also Remark 5 below, where we explain that our models cover the concave distortion as a subcase.

We say that a random variable ξ is acceptable if $u(\xi) \geq 0$. We remark that $\xi - u(\xi)$ is always acceptable. If u is coherent, then the acceptability set $\mathcal{A} = \{\xi \mid u(\xi) \geq 0\}$ is a convex cone. The continuity assumption allows one to apply convex duality theory and leads to the following representation theorem:

Theorem 1. *If u is coherent, there exists a convex closed set $\mathcal{S} \subset L^1$ (with L^1 being the vector space of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ that are integrable), consisting of probability measures, absolutely continuous with respect to $\mathbb{P}$, such that for all $\xi \in L^\infty$:*

$$u(\xi) = \inf_{\mathbb{Q} \in \mathcal{S}} \mathbb{E}_{\mathbb{Q}}[\xi].$$

Conversely, each such set $\mathcal{S}$ defines a coherent utility function.

Remark 1. *We identify an absolutely-continuous probability measure $\mathbb{Q} \ll \mathbb{P}$ with its Radon–Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$.*

Remark 2. *Replacing the true or physical probability $\mathbb{P}$ by other measures is a practice that is well known in insurance. Standard techniques that can be described are, e.g., "tilting" and increasing or decreasing ages in*

life insurance contracts. To calculate premia, more weight is then given to unfavorable events, and favorable events get less weight. The existence of such a set $\mathcal{S}$ says that the change in probability is done in a systematic and consistent way.

Remark 3. *We can show that the upward sequential continuity in the definition of coherent utility functions is equivalent to the existence of a minimizer $\mathbb{Q}_0 \in \mathcal{S}$. This equivalence follows from a deep mathematical result in functional analysis (the theorem of R.James), Delbaen (2011). That the inf can be replaced by min simplifies some of the proofs. We call this property the weak compactness property of $\mathcal{S}$.*

Definition 2. *We say that two random variables ξ, η are commonotonic if there exist a random variable ζ, as well as two non-decreasing functions $f, g \colon \mathbb{R} \to \mathbb{R}$ such that $\xi = f(\zeta)$ and $\eta = g(\zeta)$.*

Remark 4. *Loosely speaking, two risks are commonotonic if they are bets on the same event. Indeed, ξ and η being nondecreasing functions of ζ, neither of them is a hedge against the other, as both risks ineluctably move in the same direction. It is a non-trivial exercise to show that we can always take $\zeta = \xi + \eta$ (see Delbaen 2011, p. 20 for a proof of this result).*

Definition 3. *We say that $u \colon L^\infty \to \mathbb{R}$ is commonotonic if for each couple ξ, η of commonotonic random variables, we have $u(\xi + \eta) = u(\xi) + u(\eta)$.*

Remark 5. *Commonotonic concave monetary utility functions are positively homogeneous and hence coherent. In general, coherent utility functions satisfy $u(\xi + \eta) \geq u(\xi) + u(\eta)$. That means that by diversifying, the risk-adjusted valuation gets better. Commonotonic random variables form the opposite of diversification, as explained in Remark 4. Roughly speaking: what is worse for ξ is worse for η. The commonotonicity of u can therefore be seen as a translation of the rule: if there is no diversification, there is also no gain in putting these claims together. Including commonotonicity as an economic principle is desirable for the purpose of premium calculation or risk measurement in insurance, and many actuarial models are built on the assumption that the premium principles are commonotonic. For instance, the popular class of concave distortions mentioned above is comprised of commonotonic utilities; hence, they can be seen as particular examples fitting in the framework we develop below.*

2. Description of the Model

We use the following setup. There are N agents to be insured, indexed $i = 1, \ldots, N$. There is one insurer denoted by the index 0. There is also a "super"-reinsurer, whose role will be explained later. He/she will be denoted by r. The agents have liabilities that they want to insure. The liability for agent i is $X_i \geq 0$. Without insurance, his/her position will be $-X_i$. There are different premium principles, which will be described in the examples below. The utility functions of the agents are denoted by u_i. The coherent utility function of the insurer is u_0 and, to reduce complexity, supposed to be commonotonic. The utility functions of the agents are more restrictive than the insurer's utility function u_0; meaning: when a random variable is not acceptable for the insurer, then it is not acceptable for the agents. Equivalently, we can say that acceptable elements for the agent i are acceptable for the insurer. This is a translation of the fact that the agent i feels a need for insurance and that an insurer can fulfill this need. Because of the translation property, we then have for all ξ and all $i \geq 1$:

$$u_0(\xi) \geq u_i(\xi).$$

We assume the utilities $u_i, i = 1, \ldots, N$ to be monetary utility functions as in Definition 1 and satisfying the above inequality, with no further assumptions.

The insurer also brings in an initial capital k_0. He/she will take the insurance only if he/she can obtain a better outcome. Because u_0 is coherent, the total premium, π_0, must be at least $\pi_0 = -u_0\left(-\sum_{i=1}^{N} X_i\right)$. Indeed, the insurer takes the random variable $-\sum_i X_i$ and receives π_0. The deal is

only acceptable for the insurer if $u_0(\pi_0 - \sum_i X_i) \geq 0$. We suppose that the scenario set for u_0, $\mathcal{S}_0$ is weakly compact so that there is a $\mathbb{Q}_0 \in \mathcal{S}_0$ such that:

$$\mathbb{E}_{\mathbb{Q}_0}\left[\sum_{i=1}^{N} X_i\right] = \sup_{\mathbb{Q} \in \mathcal{S}_0} \mathbb{E}_{\mathbb{Q}}\left[\sum_{i=1}^{N} X_i\right] = -u_0\left(-\sum_{i=1}^{N} X_i\right).$$

In case u_0 is commonotonic, we can (and shall) even choose $\mathbb{Q}_0$ so that for all Y commonotonic with $\sum_i X_i$ and such that the distribution of Y has no other points of increase as $\sum_i X_i$, we have $u_0(Y) = \mathbb{E}_{\mathbb{Q}_0}[Y]$. This applies for cases such as $Y = \min(R, \sum_i X_i)$ or $Y = (\sum_i X_i - R)^+$), where $R \in \mathbb{R}$.

From Delbaen (2002, 2011), using the capital allocation principle, we find that the individual "fair" premia should be:

$$\pi_i = \mathbb{E}_{\mathbb{Q}_0}[X_i]. \tag{1}$$

In Delbaen (2002, 2011) it is shown that the allocation principle can also be obtained as an application of the marginal contribution of one agent. The latter was the approach given in Deprez and Gerber (1985). The equivalence requires some technical assumptions that go beyond the scope of this paper. What this means is that under these extra assumptions, we have:

$$\pi_i = \lim_{\varepsilon \downarrow 0} (-1) \frac{u_0(-\sum_j X_j - \varepsilon X_i) - u_0(-\sum_j X_j)}{\varepsilon}$$

The equality $\pi_i = \mathbb{E}_{\mathbb{Q}_0}[X_i]$ immediately implies:

$$-\pi_i = \mathbb{E}_{\mathbb{Q}_0}[-X_i] \geq u_0(-X_i) \geq u_i(-X_i).$$

For the agent i, this is a good deal, provided that the insurance pays entirely the claim X_i at Date 1, since $-\pi_i \geq u_i(-X_i)$. What he/she pays for insurance is better than paying X_i.

In the following, we propose four different models, that is four different ways in which the premia can be fixed by the insurer in situations where there is either a government guarantee or a reinsurance possibility. There is a distinction between the fair premia and what is actually charged, which may be higher. We argue that the difference should be regarded as a contributionto the capital by the insured agents, and be remunerated by a share of the surplus. We provide conditions that the deals are acceptable by all agents and the insurer. More precisely, we consider that a deal is acceptable for the insurer whenever the utility u_0 of the profit received by the insurer exceeds k_0. We say that the deal is acceptable for agent i whenever the insurance deal generates an increase in the agent's utility u_i, as compared with the situation without insurance.

3. Model 1

In this example, we suppose that the insurer has limited liability. We take for the total premium:

$$\pi_0 = -u_0\left(-\sum_{i=1}^{N} X_i\right).$$

We distinguish several cases:

1. $\sum_i X_i > \pi_0 + k_0$. In this case, the total claim size exceeds the available capital. The excess is supposed to be covered by, for instance, the government, and this at no cost. The initial capital should then be sufficiently high to make the deal acceptable for the government. The determination of this level is beyond the contents of this paper. We denote by A the set $A = \{\sum_i X_i > \pi_0 + k_0\}$.
2. $\pi_0 \leq \sum_i X_i \leq \pi_0 + k_0$. In this case, there is no surplus, and the insurer will lose part of his/her investment. We denote by B the set $B = \{\pi_0 \leq \sum_i X_i \leq \pi_0 + k_0\}$.

3. $\pi_0 > \sum_i X_i$. In this case, there is a surplus. The insurer will keep the surplus entirely. This can be defended since the agents already "gained" from the allocation principle, which is their share when entering the insurance. Furthermore, they do not take any risk. We denote by C the set $C = \{\pi_0 > \sum_i X_i\}$.

Theorem 2. *The deal is acceptable for the insurer.*

Proof. The insurer will accept the deal if:

$$u_0((\pi_0 + k_0 - \sum_i X_i)\mathbf{1}_{B \cup C}) \geq k_0.$$

This is easily proven. By the definition of π_0, we have $u_0(\pi_0 - \sum_i X_i) = 0$; hence $u_0(\pi_0 - \sum_i X_i + k_0) = k_0$; therefore, by monotonicity, $u_0\left((\pi_0 - \sum_i X_i + k_0)^+\right) \geq k_0$. □

Remark 6. *We remark that there must be a regulator, who requires a minimum capital k_0. Otherwise, the company would choose $k_0 = 0$, take the profit in the event C, and leave the trouble to "society" in the event A. As said in the introduction, the rules used by the regulators and the implementation in models is the topic of research by Artzner and Eisele.*

4. Model 2

This is almost the same as Model 1, but this time, we require a premium for covering the excess. We also assume that the reinsurer has no default. The reinsurance premium will be calculated by the same coherent utility function, i.e., the same set $\mathcal{S}_0$. It is here that we use the commonotonicity. The retention will be denoted by R, and this results in the splitting:

$$\sum_i X_i = \left(\sum_i X_i\right) \wedge R + \left(\sum_i X_i - R\right)^+.$$

The two terms are commonotonic, and hence, the premium satisfies:

$$\pi_0 = \pi^R + \rho^R; \quad \pi^R = \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_i X_i\right) \wedge R\right]; \quad \rho^R = \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_i X_i - R\right)^+\right].$$

The retention level needs to be chosen so that the claims are fully covered, that is the available funds (given by $k_0 + \pi^R$) are not below the retention level. We define the optimal retention R as the maximal retention level that results in the full coverage of the claims:

$$R := \max\{x \mid x \leq k_0 + \pi^R\}.$$

The existence and uniqueness of R follow from an easy analysis of the function Φ defined as:

$$\mathbb{R}_+ \to \mathbb{R}_+; \ x \to x - \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_i X_i\right) \wedge x\right].$$

As $x - \mathbb{E}_{\mathbb{Q}_0}\left[(\sum_i X_i) \wedge x\right] = \mathbb{E}_{\mathbb{Q}_0}\left[(x - \sum_i X_i)^+\right] = \int_0^x \mathbb{Q}_0\left(\sum_i X_i \leq a\right) da$, the function $\Phi(x)$ is continuous, convex, strictly increasing after it leaves zero, is zero at zero, and tends to ∞ for $x \to +\infty$. Hence, the optimal retention R satisfies:

$$R = k_0 + \pi^R,$$

that is, the available capital is R, which is also the maximum the insurer has to pay out. The surplus is therefore:

$$R - \left(\sum_i X_i \right) \wedge R.$$

The agents again do not take any risk, and hence, they should not participate in the surplus. The insurer finds it a good deal if:

$$u_0 \left(R - \left(\sum_i X_i \right) \wedge R \right) \geq k_0.$$

However, the definition of R shows that $u_0 \left(R - \left(\sum_i X_i \right) \wedge R \right) = k_0$. That means there is no incentive to do business, and the insurer must get all the profit to have an equivalent outcome.

5. Model 3

This is an extension of the previous models. The agents pay a premium equal to $p_i \geq \pi_i$, where π_i is the fair premium, introduced in Equation (1). This has the advantage that the insurer can announce the premium without having to calculate the total premium necessary to cover the total losses. Of course, this procedure should lead to a premium greater than the fair premium, as calculated in the previous models. An example of such a premium calculation could be the amount $\sup_{\mathbb{Q} \in \mathcal{S}_0} \mathbb{E}_{\mathbb{Q}}[X_i] = -u_0(-X_i)$. The extra premium $p_i - \pi_i$ can be seen as a contribution of agent i to the capital; hence, agents should be entitled to a share in the surplus. All investors must be treated in the same way, and hence, the share of agent i is proportional to his/her contribution, namely $\frac{p_i - \pi_i}{\sum_j (p_j - \pi_j) + k_0}$. The investor will get a proportion $\frac{k_0}{\sum_j (p_j - \pi_j) + k_0}$. These fractions are paid out regardless of having caused a claim or not.

The retention is now defined by the relation:

$$\sum_i (p_i - \pi_i) + k_0 + \pi^R = R.$$

The existence and uniqueness of R are proven in the same way. This time, we must see whether this is a good deal for the insurer, as well as for the agents.

Theorem 3. *The deal is acceptable for the insurer.*

Proof. For the insurer, we must check the inequality:

$$u_0 \left(\frac{k_0}{k_0 + \sum_i (p_i - \pi_i)} \left(R - \sum_i X_i \right)^+ \right) \geq k_0.$$

By positive homogeneity of u_0, this is the same as:

$$u_0 \left(\left(R - \sum_i X_i \right)^+ \right) \geq k_0 + \sum_i (p_i - \pi_i).$$

As in the previous example, this follows from the definition of R, which implies:

$$u_0 \left(R - \left(\sum_i X_i \right) \wedge R \right) = R - \pi^R = k_0 + \sum_i (p_i - \pi_i).$$

□

Theorem 4. *The deal is acceptable for the insured as soon as* $p_i \leq \sup_{\mathbb{Q} \in \mathcal{S}_0} \mathbb{E}_{\mathbb{Q}}[X_i]$.

Proof. For agent i, we must check:

$$u_i\left(-p_i + \frac{p_i - \pi_i}{k_0 + \sum_j (p_j - \pi_j)} \left(R - \sum_i X_i\right)^+\right) \geq u_i(-X_i).$$

This is equivalent to:

$$u_i\left(\frac{p_i - \pi_i}{k_0 + \sum_j (p_j - \pi_j)} \left(R - \sum_i X_i\right)^+\right) \geq u_i(p_i - X_i).$$

The left-hand side is positive, whereas the right-hand side is negative, provided the premium p_i is not too big. For instance, if $p_i \leq \sup_{\mathbb{Q} \in \mathcal{S}_0} \mathbb{E}_{\mathbb{Q}}[X_i]$, we have $u_0(p_i - X_i) \leq 0$ and, hence, also $u_i(p_i - X_i) \leq 0$. $\square$

Remark 7. *In any case, the agent i will not pay a premium p_i that is bigger than $-u_i(-X_i)$. Paying a higher premium and counting on surplus participation is not realistic since the surplus share also depends on the claims incurred through the other agents.*

6. Model 4

We continue the building of stepwise more complicated models. We suppose that there are two insurers. The first one has a utility function u_0, described by the scenario set $\mathcal{S}_0$. This insurer acts as the direct insurer. The second insurer acts as a reinsurer with utility function u_r described by the scenario-set $\mathcal{S}_r$. For a claim $0 \leq \xi$, the reinsurer would charge a premium $\sup_{\mathbb{Q} \in \mathcal{S}_r} \mathbb{E}_{\mathbb{Q}}[\xi]$.

Both utility functions u_0 and u_r are supposed to be commonotone. Their scenario-sets are therefore determined as cores of convex games, say $v_0 \geq v_r$. For simplicity, we suppose that both scenario-sets $\mathcal{S}_0, \mathcal{S}_r$ are weakly compact. We can therefore suppose that there are elements $\mathbb{Q}_0 \in \mathcal{S}_0, \mathbb{Q}_r \in \mathcal{S}_r$, such that for all a:

$$\sup_{\mathbb{Q} \in \mathcal{S}_r} \mathbb{E}_{\mathbb{Q}}\left[\sum_j X_j\right] = \mathbb{E}_{\mathbb{Q}_r}\left[\sum_j X_j\right]$$

$$\sup_{\mathbb{Q} \in \mathcal{S}_r} \mathbb{E}_{\mathbb{Q}}\left[\left(\sum_j X_j\right) \wedge a\right] = \mathbb{E}_{\mathbb{Q}_r}\left[\left(\sum_j X_j\right) \wedge a\right]$$

$$\sup_{\mathbb{Q} \in \mathcal{S}_r} \mathbb{E}_{\mathbb{Q}}\left[\left(\sum_j X_j - a\right)^+\right] = \mathbb{E}_{\mathbb{Q}_r}\left[\left(\sum_j X_j - a\right)^+\right]$$

$$\sup_{\mathbb{Q} \in \mathcal{S}_0} \mathbb{E}_{\mathbb{Q}}\left[\sum_j X_j\right] = \mathbb{E}_{\mathbb{Q}_0}\left[\sum_j X_j\right]$$

$$\sup_{\mathbb{Q} \in \mathcal{S}_0} \mathbb{E}_{\mathbb{Q}}\left[\left(\sum_j X_j\right) \wedge a\right] = \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_j X_j\right) \wedge a\right]$$

$$\sup_{\mathbb{Q} \in \mathcal{S}_0} \mathbb{E}_{\mathbb{Q}}\left[\left(\sum_j X_j - a\right)^+\right] = \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_j X_j - a\right)^+\right].$$

The reinsurer is supposed to be default-free (it could be a government institution or a solidarity fund of the insurance industry). We suppose that the agents have access to this reinsurer. We assume that:

$$u_i \leq u_r \leq u_0 \text{ for all } i,$$

in particular $\mathcal{S}_0 \subset \mathcal{S}_r$. This reflects the fact that on the one hand, the agents are more risk-adverse than the insurers, and on the other hand, the reinsurer being default free, the premia for the reinsurer are higher than for the direct insurer.

The actions of the direct insurer are subject to the rule that after reinsurance, the claims must be covered completely. The direct insurer will therefore take a reinsurance with a retention R, to be determined later. For the retention level R, the cost of the reinsurance is given by (using the commonotonicity of u_r):

$$\rho^R := -u_r \left(-\left(\sum_j X_j - R \right)^+ \right) = \mathbb{E}_{\mathbb{Q}_r} \left[\left(\sum_j X_j - R \right)^+ \right].$$

To get full coverage of their risks without the input of any capital (that is, $k_0 = R = 0$), the agents should pay a total premium:

$$\rho^0 = -u_r \left(-\sum_j X_j \right) = \mathbb{E}_{\mathbb{Q}_r} \left[\sum_j X_j \right],$$

as the direct insurer can transfer the totality of the claims to the reinsurer at the cost ρ^0. Using the capital allocation principle, we deduce that the premium for such a coverage to be paid by agent j is:

$$\pi_j^r := \mathbb{E}_{\mathbb{Q}_r} \left[X_j \right].$$

For the moment, we consider the amounts π_j^r as the premium-input of the agent j; if they pay higher premia, p_j, the differences, $p_j - \pi_j^r$, are regarded as capital-input. In practice however, the direct insurer brings in a capital $k_0 > 0$ (again, we consider this to be fixed by a regulator), and hence, the retention R will not be zero. This means that whenever $R > 0$, the premia π_j^r that we are considering are not exactly the fair premia (see also Remark 8 below).

As the insurer participates with capital k_0, we must determine the retention level R; as before, we consider R to be the maximal retention level that results in full coverage. The full coverage condition is now given by: $R \leq k_0 + \sum_j p_j - \rho^R$, so that the optimal retention is:

$$R = \max\{x \mid x \leq k_0 + \sum_j p_j - \rho^R\}.$$

The full coverage condition can be written as: $R - \mathbb{E}_{\mathbb{Q}_r} \left[\left(\sum_j X_j \right) \wedge R \right] \leq k_0 + \sum_j (p_j - \pi_j^r)$.

Therefore, the existence and unicity of optimal retention comes using identical arguments as before for Φ, but this time for the function Ψ, defined as:

$$\mathbb{R}_+ \to \mathbb{R}_+; \ x \to x - \mathbb{E}_{\mathbb{Q}_r} \left[\left(\sum_j X_j \right) \wedge x \right].$$

From now on, we always assume R to be the optimal retention, that is: $R = k_0 + \sum_j p_j - \rho^R$, or, alternatively:

$$\Psi(R) = k_0 + \sum_j (p_j - \pi_j^r).$$

In Remark 9, we analyze the dependence of the optimal reserve on the level of the capital.

With the capital input of the insurer being k_0 and the retention level R, the outcome (i.e., surplus) at the end of the contract is:

$$S := k_0 + \sum_j p_j - \rho^R - \left(\sum_j X_j \right) \wedge R \geq 0.$$

We denote by $\lambda_0 = \frac{k_0}{k_0 + \sum_j (p_j - \pi_j^r)}$ the proportion of the surplus that the direct insurer is keeping and by $\lambda_i = \frac{p_i - \pi_i^r}{k_0 + \sum_j (p_j - \pi_j^r)}$ the proportion of the surplus that agent i is keeping.

We now prove that this procedure of surplus sharing is beneficial for the insurer and for the insured. Let us fist check the utility for the direct insurer.

Theorem 5. *The deal is acceptable for the direct insurer.*

Proof. For the direct insurer, the deal is acceptable if and only if the utility of his/her share of the surplus is not less that k_0, that is $u_0(\lambda_0 S) \geq k_0$. The utility of his share is given by:

$$u_0(\lambda_0 S) = \lambda_0 \left(k_0 + \sum_j p_j - \rho^R - \mathbb{E}_{\mathbb{Q}_0} \left[\left(\sum_j X_j \right) \wedge R \right] \right).$$

This quantity is bigger than k_0 if and only if:

$$-\mathbb{E}_{\mathbb{Q}_0} \left[\left(\sum_j X_j \right) \wedge R \right] - \mathbb{E}_{\mathbb{Q}_r} \left[\left(\sum_j X_j - R \right)^+ \right] \geq \sum_j (-\pi_j^r) = -\mathbb{E}_{\mathbb{Q}_r} \left[\sum_j X_j \right] = u_r \left(-\sum_j X_j \right).$$

The utility functions u_0 and u_r are commonotone, and therefore, we have that:

$$u_0 \left(-\left(\sum_j X_j \right) \wedge R \right) = -\mathbb{E}_{\mathbb{Q}_0} \left[\left(\sum_j X_j \right) \wedge R \right]$$

$$u_r \left(-\left(\sum_j X_j - R \right)^+ \right) = -\mathbb{E}_{\mathbb{Q}_r} \left[\left(\sum_j X_j - R \right)^+ \right].$$

Therefore, the inequality is obvious, given $u_0 \geq u_r$. □

Theorem 6. *The deal is acceptable for the insured as soon as $p_i \leq \sup_{\mathbb{Q} \in \mathcal{S}_r} \mathbb{E}_{\mathbb{Q}}[X_i]$.*

Proof. To check the advantage for the insured i, we must show that $u_i(\lambda_i S - p_i) \geq u_i(-X_i)$, that is:

$$u_i(\lambda_i S) \geq p_i + u_i(-X_i)$$

The surplus being nonnegative, the left-hand side is nonnegative. The right side is negative provided the premium p_i is not too big. For instance, if $p_i \leq \sup_{\mathbb{Q} \in \mathcal{S}_r} \mathbb{E}_{\mathbb{Q}}[X_i]$, we have $u_r (p_i - X_i) \leq 0$ and, hence, also $u_i (p_i - X_i) \leq 0$. □

Remark 8. *For each agent, the cost p_i was split into two parts: premium-input π_i^r and capital-input $p_i - \pi_i^r$, with all premia-input summing up to $\mathbb{E}_{\mathbb{Q}_r}[\sum_j X_j] = -u_r \left(-\sum_j X_j \right)$. This is higher than the total required premium, which should be the cost of the reinsurance contract plus the cost of the direct insurance of the claims up to the retention level, given as:*

$$-u_r\left(-\left(\sum_j X_j - R\right)^+\right) - u_0\left(-\left(\sum_j X_j\right)\wedge R\right)$$

$$= \mathbb{E}_{\mathbb{Q}_r}\left[\left(\sum_j X_j - R\right)^+\right] + \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_j X_j\right)\wedge R\right].$$

Indeed, from our assumptions on the utility functions, we have:

$$\mathbb{E}_{\mathbb{Q}_r}\left[\left(\sum_j X_j - R\right)^+\right] + \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_j X_j\right)\wedge R\right] \leq \mathbb{E}_{\mathbb{Q}_r}\left[\sum_j X_j\right].$$

Probably, a more fair way to split the cost p_i is to consider:

$$\pi_i = \mathbb{E}_{\mathbb{Q}_r}\left[X_i \mathbf{1}_{\{\sum_j X_j > R\}}\right] + \mathbb{E}_{\mathbb{Q}_0}\left[X_i \mathbf{1}_{\{\sum_j X_j \leq R\}}\right]$$

as being the premium-input of agent i and the remaining $p_i - \pi_i$ as being the capital-input of agent i (provided this is positive). This different way of splitting the cost does not affect (of course) the way of computing the optimal retention level, nor the surplus S available after payment of all claims. It is only meant to provide an alternative rule for sharing the surplus: now, $\frac{k_0}{k_0 + \sum_j(p_j - \pi_j)}$ is the proportion of the surplus that the direct insurer is keeping and $\frac{p_i - \pi_i}{k_0 + \sum_j(p_j - \pi_j)}$ the proportion of the surplus that agent i is keeping. Nevertheless, for such allocations of the surplus, we did not find a simple condition for the premia p_i that ensures that the deals are acceptable for the agents, while respecting $p_i \geq \pi_i$. By a simple rule, we mean a condition on all p_i that does not involve the agent's utility directly.

However, the reader can easily check that for the direct insurer, any allocation among the insured agents of the quantity $\mathbb{E}_{\mathbb{Q}_r}\left[\left(\sum_j X_j - R\right)^+\right] + \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_j X_j\right)\wedge R\right]$ leaves the insurer with an outcome that in u_0 utility is equivalent to the initial capital k_0. This is even independent of the retention limit.

Remark 9. *We will now show that with increasing capital k_0, the utility for the direct insurer goes up. Let us denote by $R(k_0)$ the solution of:*

$$R - \mathbb{E}_{\mathbb{Q}_r}\left[\left(\sum_j X_j\right)\wedge R\right] = k_0 + \sum_j(p_j - \pi_j^r).$$

We recall that the function $\Psi(R) = R - \mathbb{E}_{\mathbb{Q}_r}\left[\left(\sum_j X_j\right)\wedge R\right]$ satisfies $\Psi(0) = 0$ and is continuous with $\Psi'(R) = \mathbb{Q}\left(\sum_j X_j \leq R\right)$. Hence, it is strictly increasing after it leaves zero and tends to ∞ when R tends to ∞. Furthermore, it is convex. These properties ensure that Ψ^{-1} is well defined on $(0, \infty)$ and is continuous, strictly increasing, and concave.

Furthermore, we recall, that we consider that all $p_j \geq \pi_j^r$ so that $\sum_j(p_j - \pi_j^r) \geq 0$. We can consider $k_0 > 0$, to ensure the positivity of the quantity $k_0 + \sum_j(p_j - \pi_j^r)$. The implicit function theorem then shows that $\frac{dR}{dk_0} \geq 0$.

The utility of the insurer's share of the surplus $u_0(\lambda_0 S)$ can be written as:

$$\lambda_0 \left(k_0 + \sum_j p_j - \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_j X_j\right) \wedge R\right] - \mathbb{E}_{\mathbb{Q}_r}\left[\left(\sum_j X_j - R\right)^+\right]\right)$$

$$= \frac{k_0}{k_0 + \sum_j(p_j - \pi_j^r)}\left(k_0 + \sum_j(p_j - \pi_j^r) - \mathbb{E}_{\mathbb{Q}_r}\left[\left(\sum_j X_j\right) \wedge R\right] - \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_j X_j\right) \wedge R\right]\right)$$

$$= \frac{k_0}{k_0 + \sum_j(p_j - \pi_j^r)}\left(R - \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_j X_j\right) \wedge R\right]\right)$$

$$= \frac{k_0}{k_0 + \sum_j(p_j - \pi_j^r)}\mathbb{E}_{\mathbb{Q}_0}\left[\left(R - \sum_j X_j\right)^+\right].$$

The first factor is clearly increasing in k_0, and the second factor is increasing in R, hence also in k_0. More important is the difference between this utility and the initial capital k_0:

$$u_0(\lambda_0 S) - k_0 = \frac{k_0}{k_0 + \sum_j(p_j - \pi_j^r)}\left(R - \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_j X_j\right) \wedge R\right]\right) - k_0$$

$$= k_0\left(\frac{R - \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_j X_j\right) \wedge R\right]}{k_0 + \sum_j p_j - \mathbb{E}_{\mathbb{Q}_r}\left[\sum_j X_j\right]} - 1\right)$$

$$= k_0\left(\frac{\mathbb{E}_{\mathbb{Q}_r}\left[\left(\sum_j X_j\right) \wedge R\right] - \mathbb{E}_{\mathbb{Q}_0}\left[\left(\sum_j X_j\right) \wedge R\right]}{k_0 + \sum_j p_j - \mathbb{E}_{\mathbb{Q}_r}\left[\sum_j X_j\right]}\right).$$

This shows that when $k_0 \to \infty$, the extra return tends to $\mathbb{E}_{\mathbb{Q}_r}\left[\sum_j X_j\right] - \mathbb{E}_{\mathbb{Q}_0}\left[\sum_j X_j\right]$, but also shows that the return on the initial capital tends to zero.

7. Discussion of the Models

There are different shortcomings of the models used. The premia paid by the agents are augmented by the administration or handling costs. Here, we might argue that in the case of a claim, the agents incur these costs themselves. In the handling costs, there is also included the commission paid out to the intermediaries, the brokers. These commissions should not be too high since otherwise, the agents could keep their liabilities or take insurance only for the larger part of the claims, the so-called tail of the distribution.

We believe that in general, the utility functions of the insured agents should not be supposed positively homogeneous. They should be concave. For this reason, we did not make any assumption on the utility function of an agent, except that it is more restrictive than that of an insurer, which seemed to us to be a realistic property. The inequalities $u_i \leq u_r \leq u_0$ can be verified in case u_0 and u_r are positively homogeneous and u_i is only concave. However, the inequality implies some geometric restrictions on the acceptability set of agent i. For instance, the function u_i cannot be differentiable at $0 \in L^\infty$. Note also that whenever commonotonicity is used, one must suppose that the utility functions are positively homogeneous, as this is a consequence of commonotonicity, Delbaen (2011). In the current paper, we only used coherence and commonotonicity for the utility functions of the insurer and the reinsurer, but these properties were not used for the insured (where we used only the monotonicity of the utility functions and the inequalities implied by the more restrictive utilities).

We supposed the presence of a reinsurer who is default free. In some cases, a government can provide a guarantee, but in general, we must include the possibility of default of the reinsurer.

Some countries require guarantees from the reinsurer either under the form of deposits or under the form of letters of credit issued by "bona fide" financial institutions.

In the case of liability insurance, there might be a rule that in case of the default of the insurer and of the reinsurer, the agent is not liable for the remaining losses. Especially in high-end insurance contracts, such rules and exceptions make the modeling extremely difficult.

In our models, there is a reward for those who take the risk. This is in contradiction with some life insurance practices, where only the amount of the total premium is important.

References

Artzner, Philippe, and Karl-Theodor Eisele. 2018. Oral communication. Strasbourg.

Delbaen, Freddy. 1974. Convex Games and Extreme Points. *Journal of Mathematical Analysis and Applications* 45: 210–33. [CrossRef]

Delbaen, Freddy. 2002. *Coherent Risk Measures. Lectures Given at The Cattedra Galileiana at the Scuola Normale Superiore Di Pisa, March 2000*. Pisa: Scuola Normale Superiore di Pisa.

Delbaen, Freddy. 2011. *Monetary Utility Functions. Lectures Held in 2008 and Published in the Series "Lecture Notes of the University of Osaka"*. Osaka: Osaka University.

Denneberg, Dieter. 1989. Verzerrte Wahrscheinlichkeiten in der. Versicherungsmathematik, quantilabhängige Prämienprinzipen. In *Mathematik-Arbeitspapiere*. Bremen: Universität Bremen, vol. 34.

Denneberg, Dieter, and Dmitri Kaplan. 1998. Die Methode der verzerrten Wahrscheinlichkeiten in der Lebens- und Rentenversicherung. *Blätter DGVFM* 23: 473–87. [CrossRef]

Deprez, Olivier, and Hans U. Gerber. 1985. On convex principles of premium calculation. *Insurance: Mathematics and Economics* 4: 179–89. [CrossRef]

Shapley, Lloyd S. 1971. Cores of Convex Games. *International Journal of Game Theory* 40: 214–25. [CrossRef]

Schmeidler, David. 1986. Integral Representation without Additivity. *Proceedings of the American Mathematical Society* 97: 255–61. [CrossRef]

Wang, Tan. 2002. A Class of Multi-Prior Preferences. Working Paper. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.197.9509&rep=rep1&type=pdf (accessed on 20 December 2018).

Yaari, Menahem E. 1987. The dual theory of choice under risk. *Econometrica* 55: 95–115. [CrossRef]